Universidad Popular del Cesar

Minería de datos y Patrones

Version 2024-I

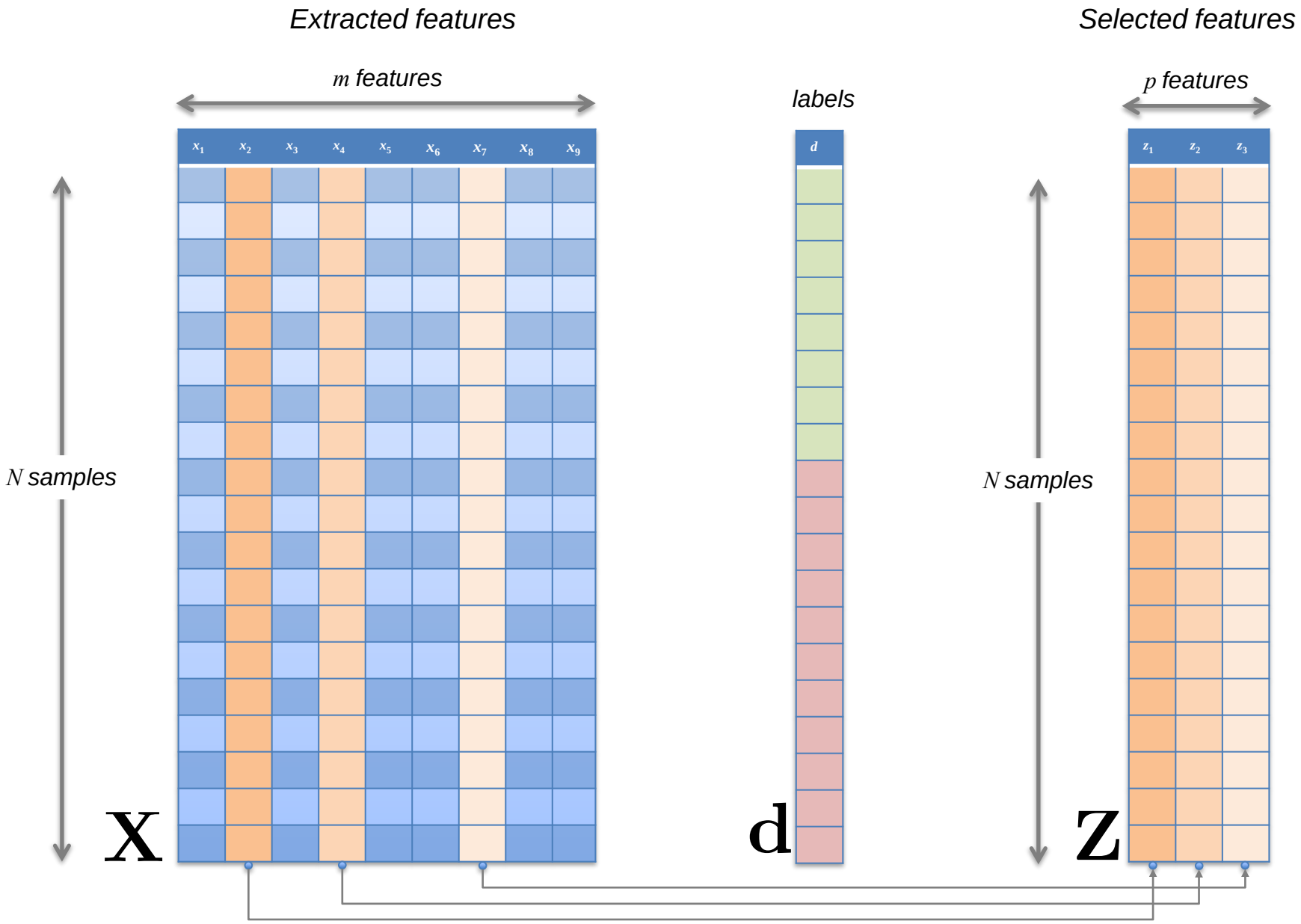
Feature Selection
SFS - Sequential Forward Selection

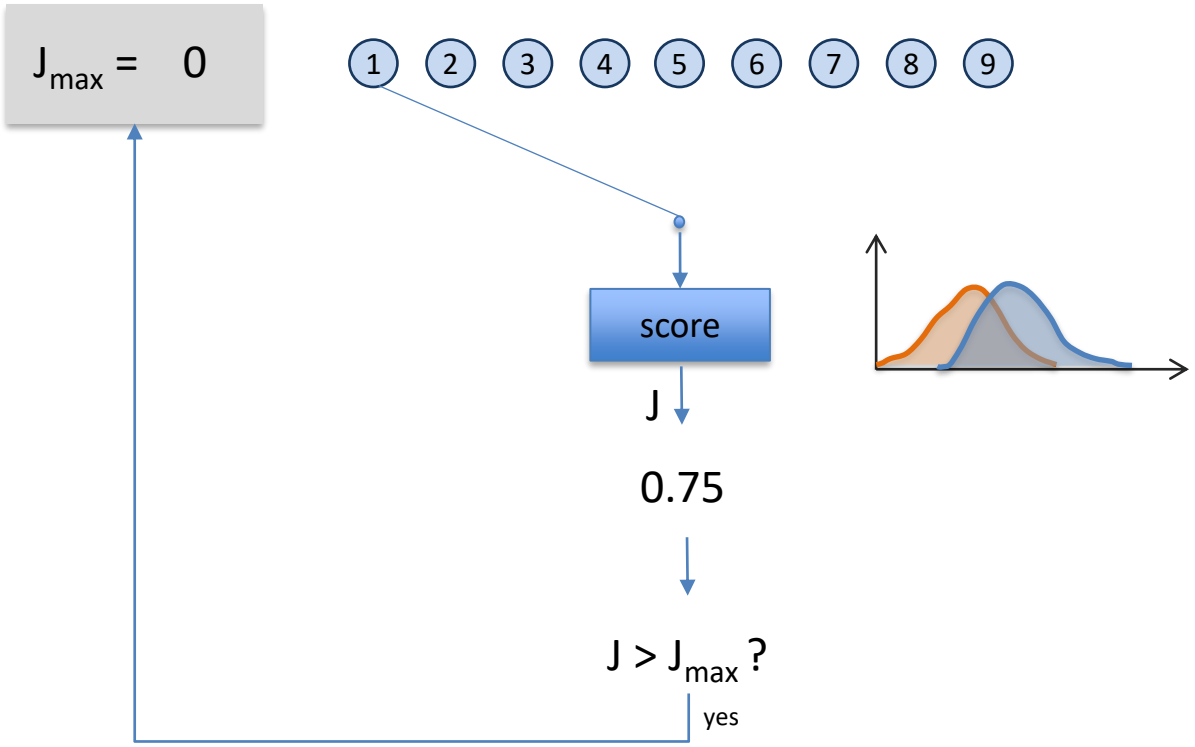
[Capítulo 3]

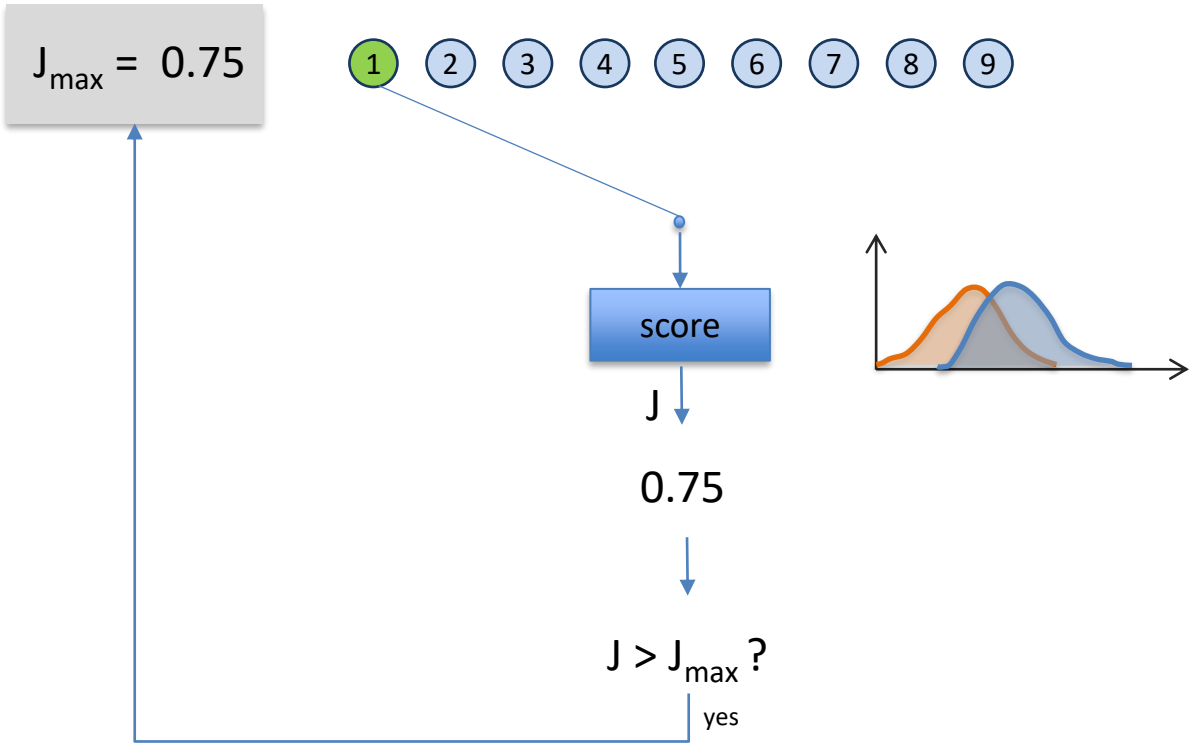
Dr. José Ramón Iglesias
DSP-ASIC BUILDER GROUP
Director Semillero TRIAC
Ingenieria Electronica
Universidad Popular del Cesar

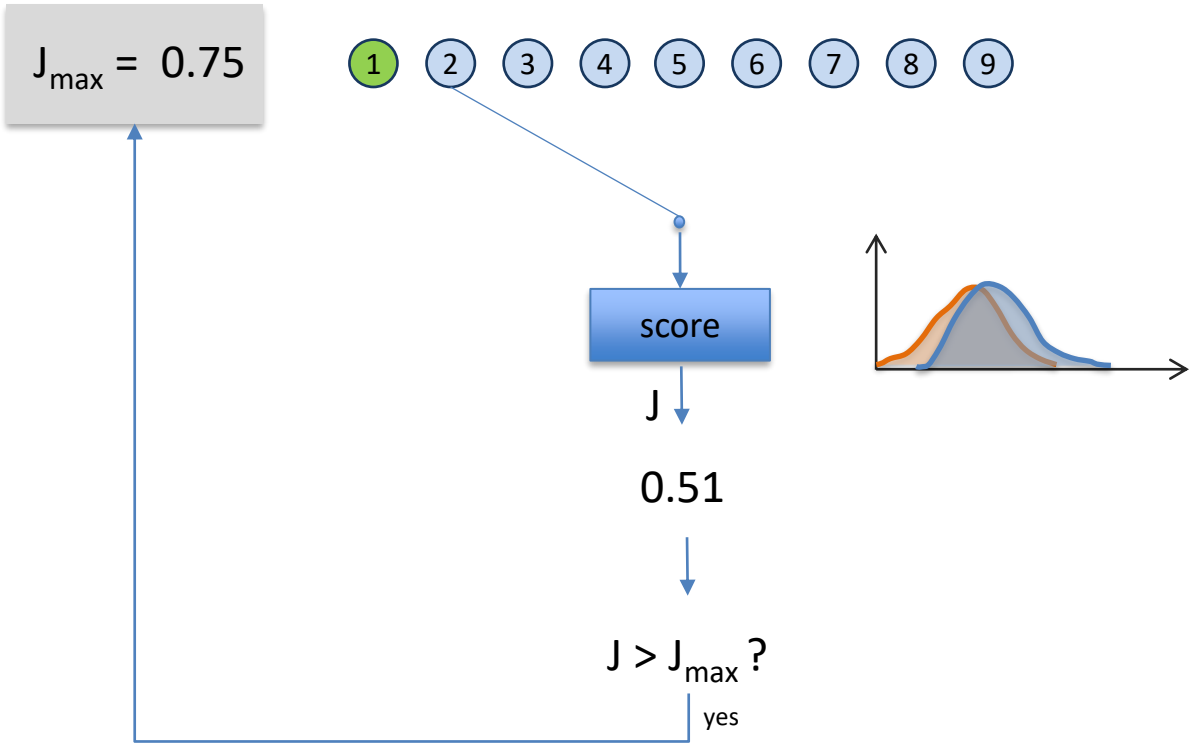
Example:

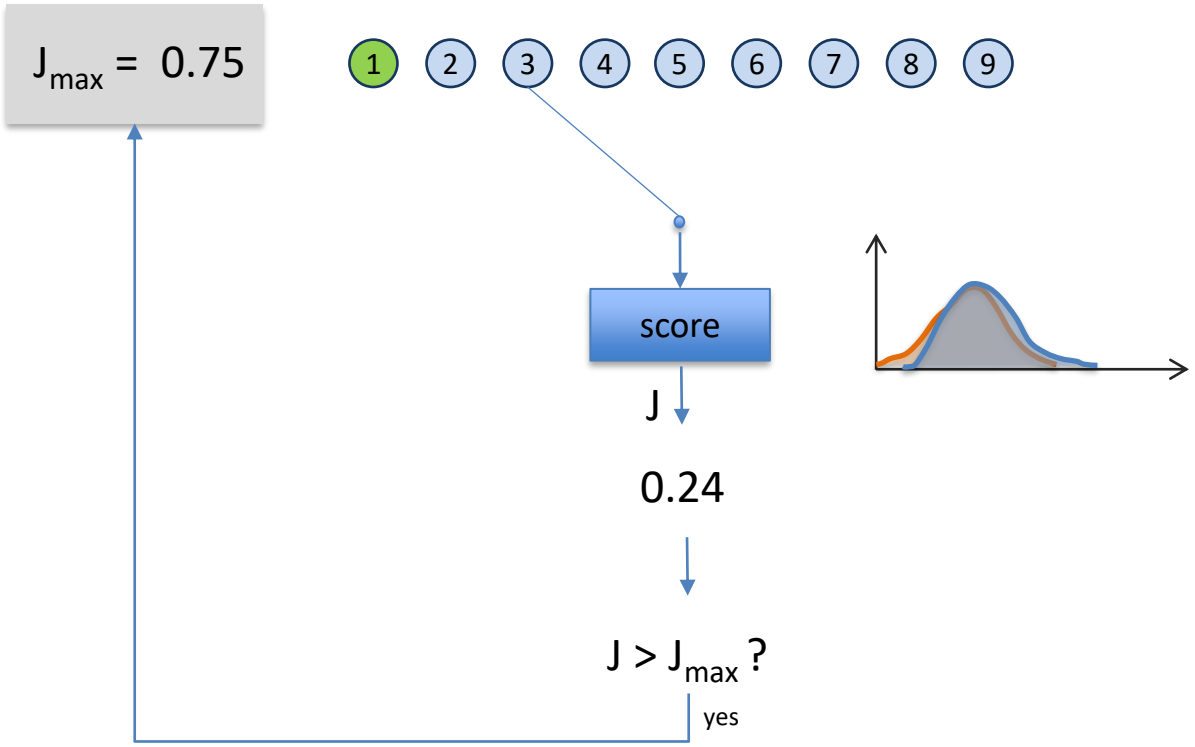
From 9 features, select 3

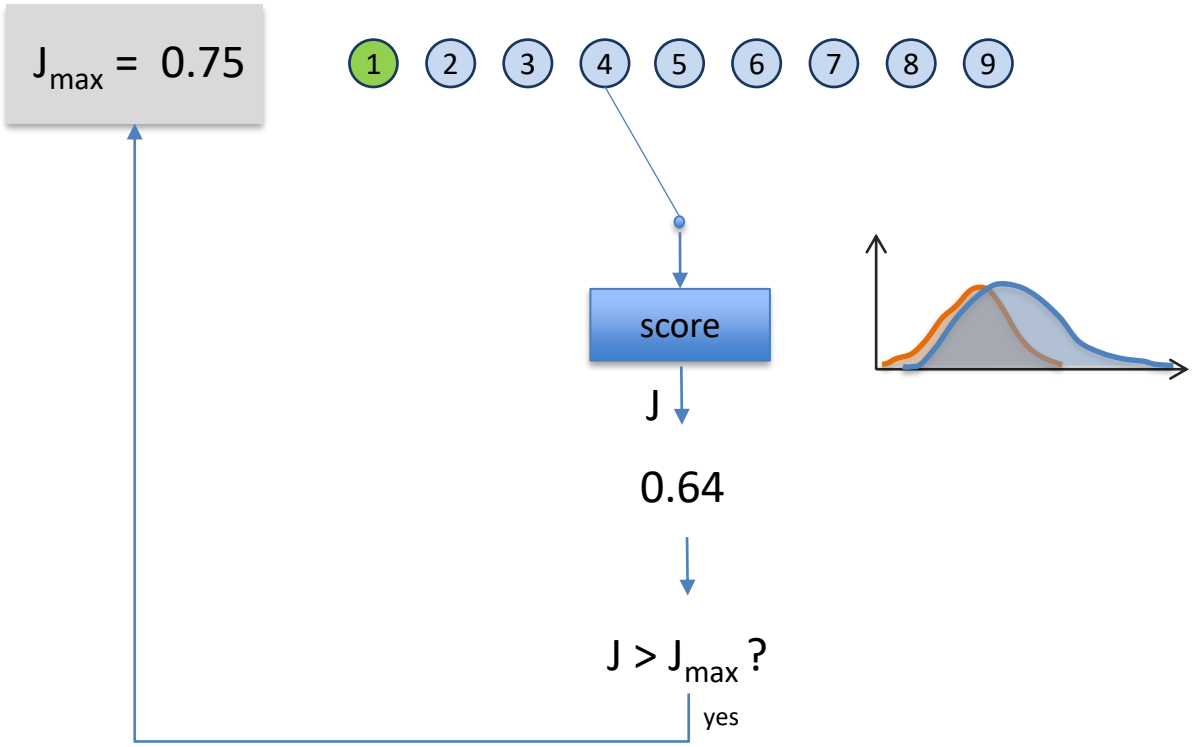

Extracted features
Selected features
m features
p features
labels
N samples
N samples
x1 x2 x3 x4 x5 x6 x7 x8 x9
d
z1 z2 z3
X
d
Z

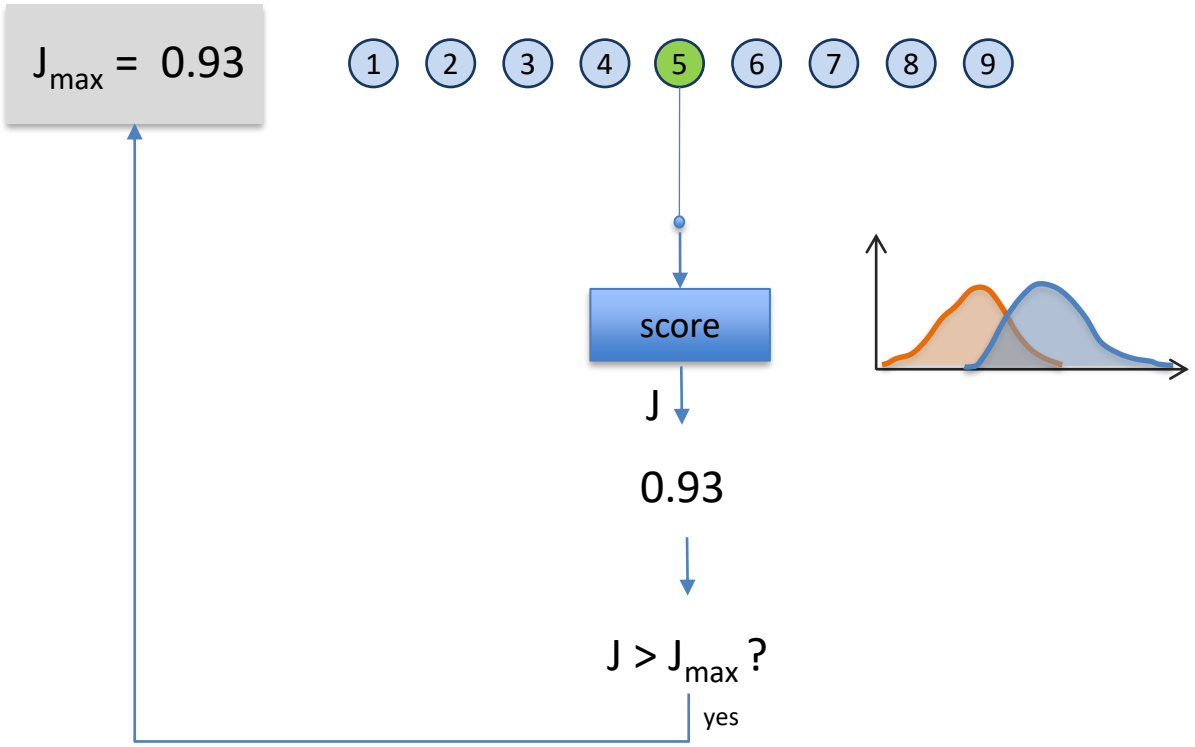

J_{max} = 0
1
2
3
4
5
6
7
8
9
score
J
0.75
$J > J_{max}$?
yes

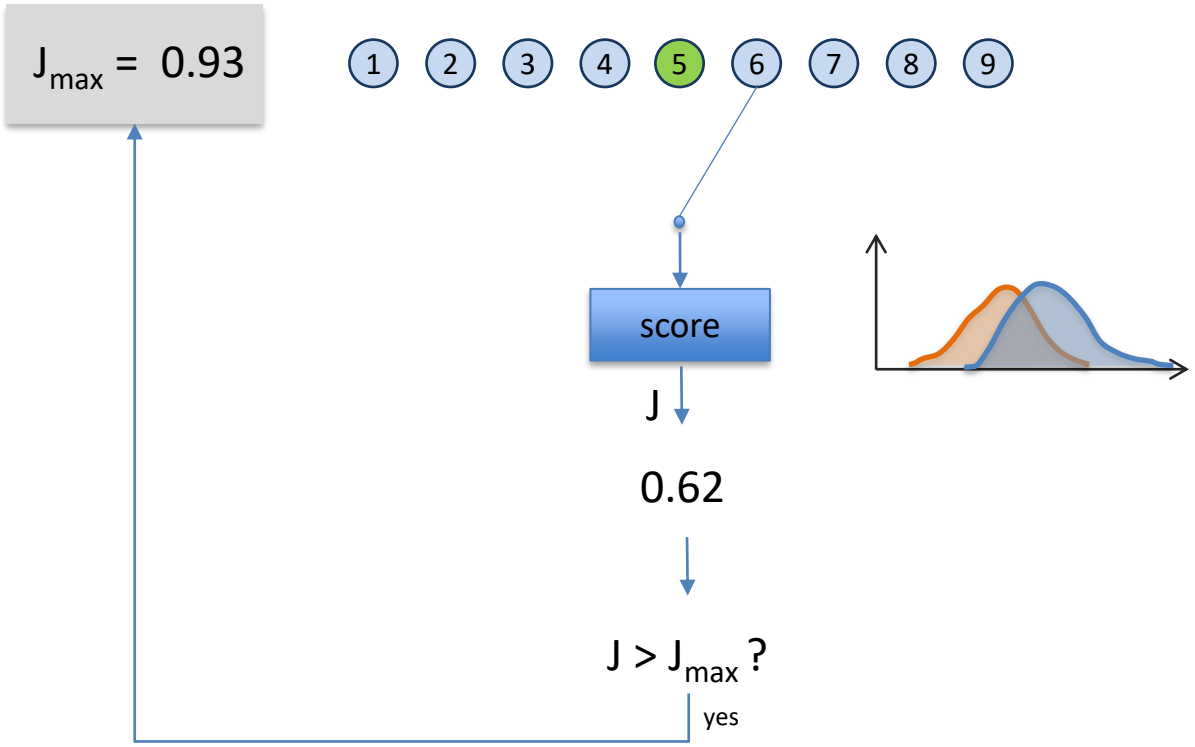

J_{max} = 0.75
1
2
3
4
5
6
7
8
9
score
J
0.75
$J > J_{max}$?
yes

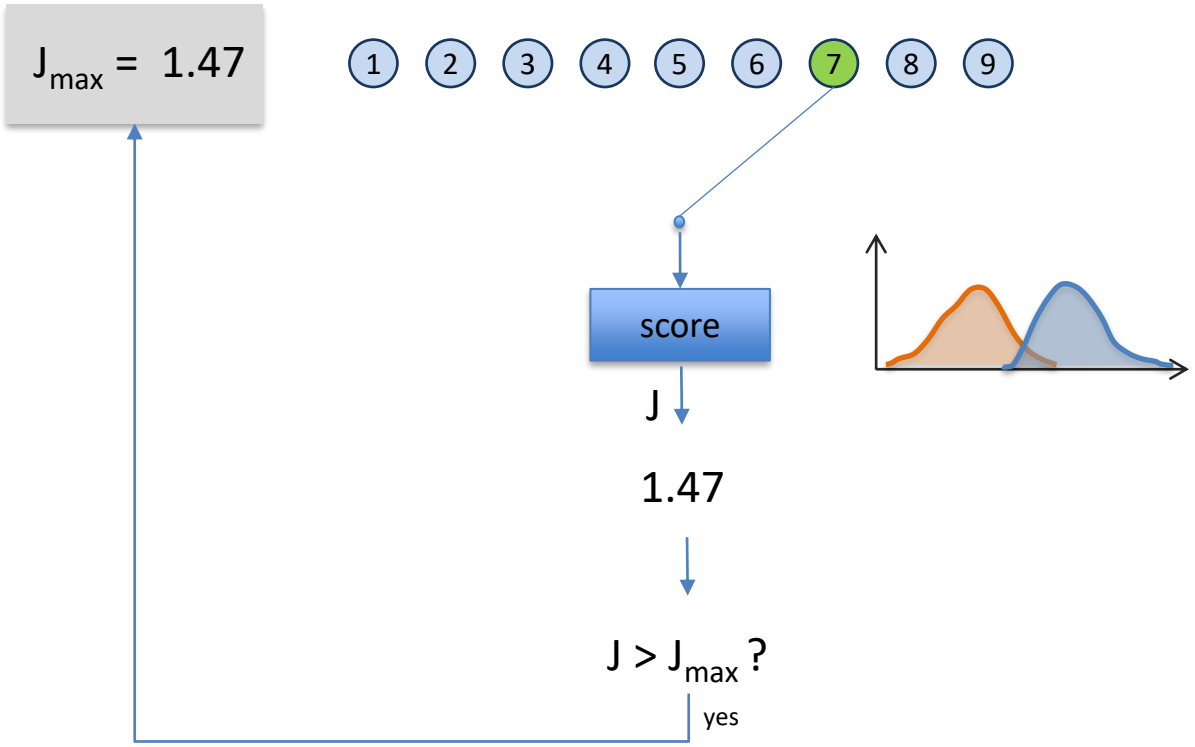

J_{max} = 0.75
1
2
3
4
5
6
7
8
9
score
J
0.51
$J > J_{max}$?
yes

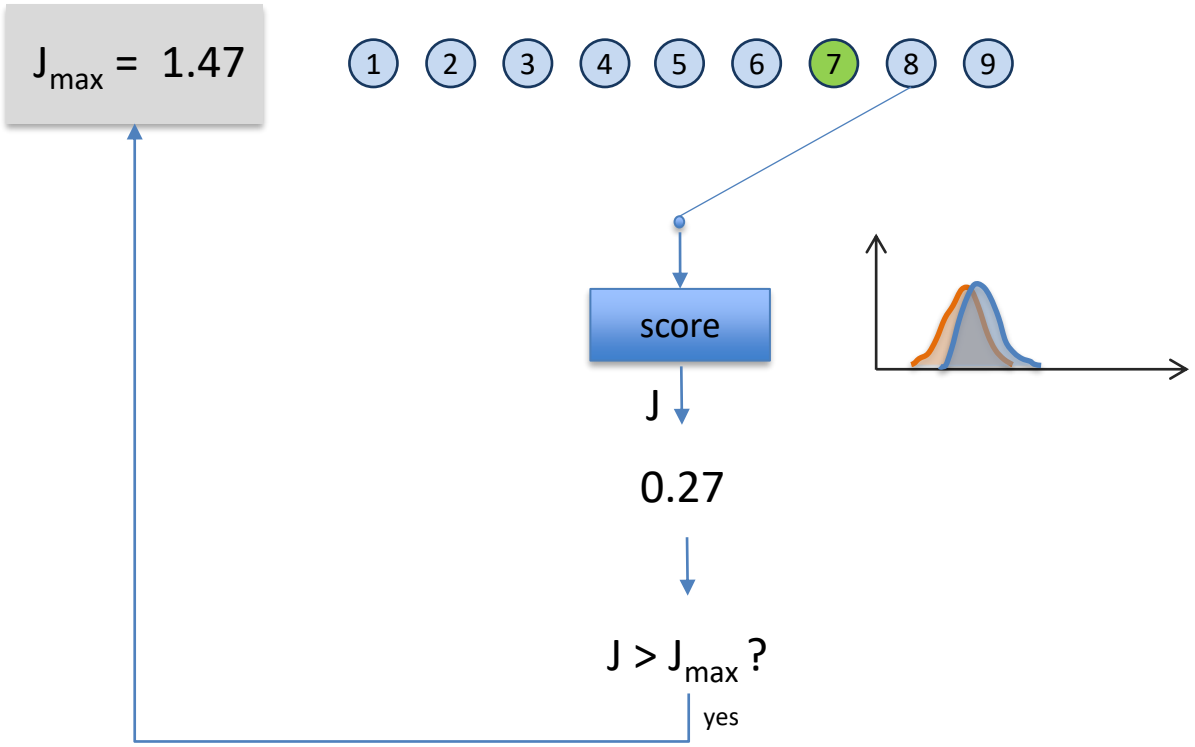

J_{max} = 0.75
1
2
3
4
5
6
7
8
9
score
J
0.24
J > J_{max} ?
yes

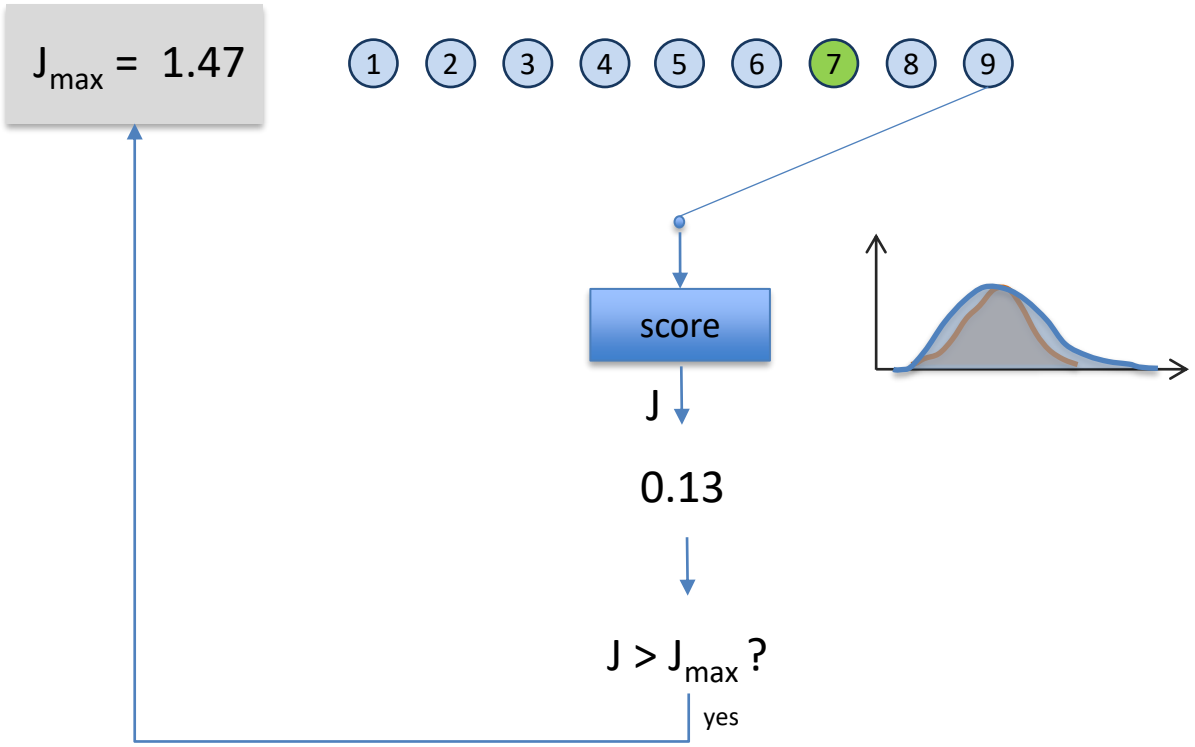

J_{max} = 0.75
1
2
3
4
5
6
7
8
9
score
J
0.64
$J > J_{max}$?
yes


J_{max} = 0.93
1
2
3
4
5
6
7
8
9
score
J
0.93
J > J_{max} ?
yes

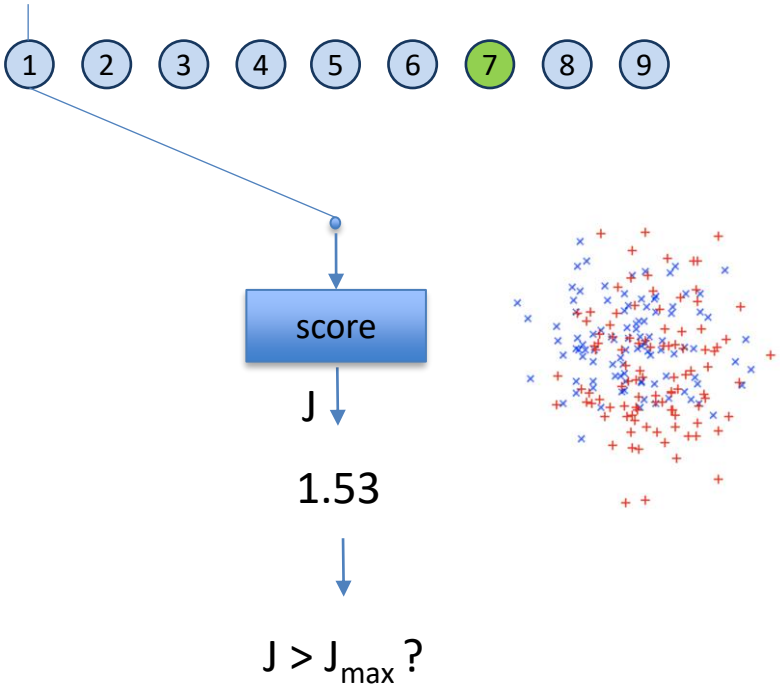

J_{max} = 0.93
1
2
3
4
5
6
7
8
9
score
J
0.62
$J > J_{max}$?
yes

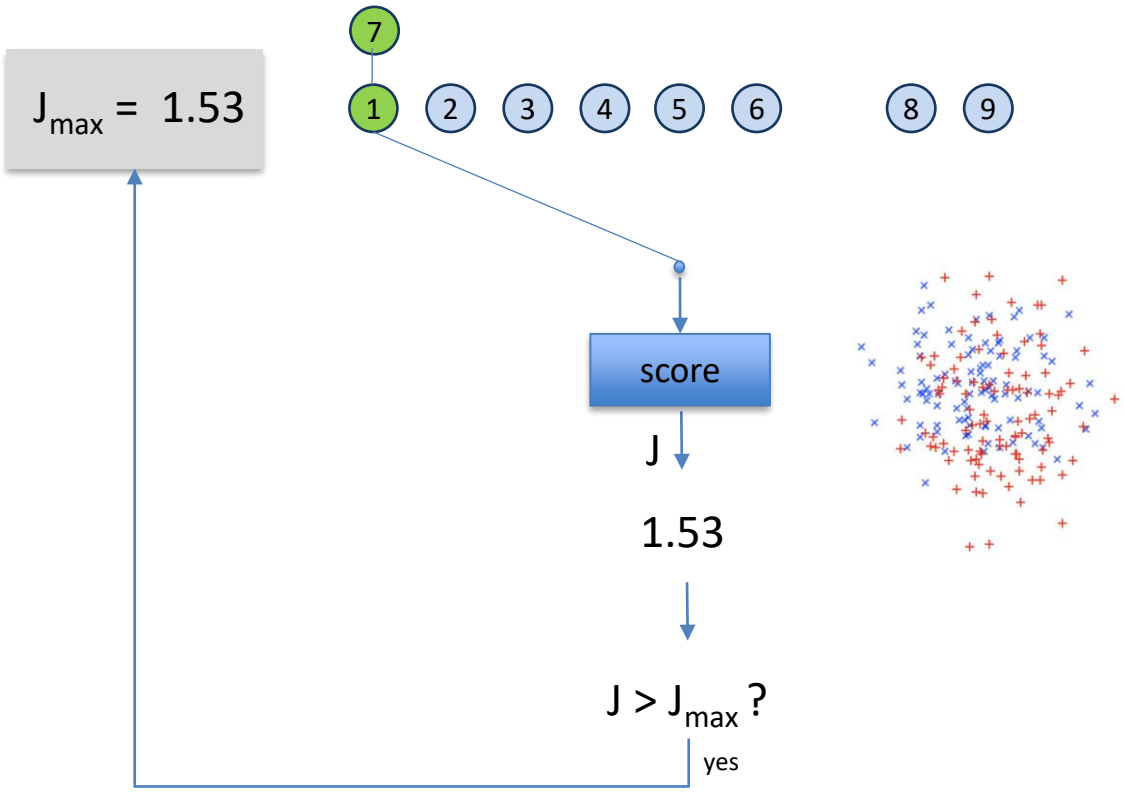

J_{max} = 1.47
1
2
3
4
5
6
7
8
9
score
J
1.47
J > J_{max} ?
yes

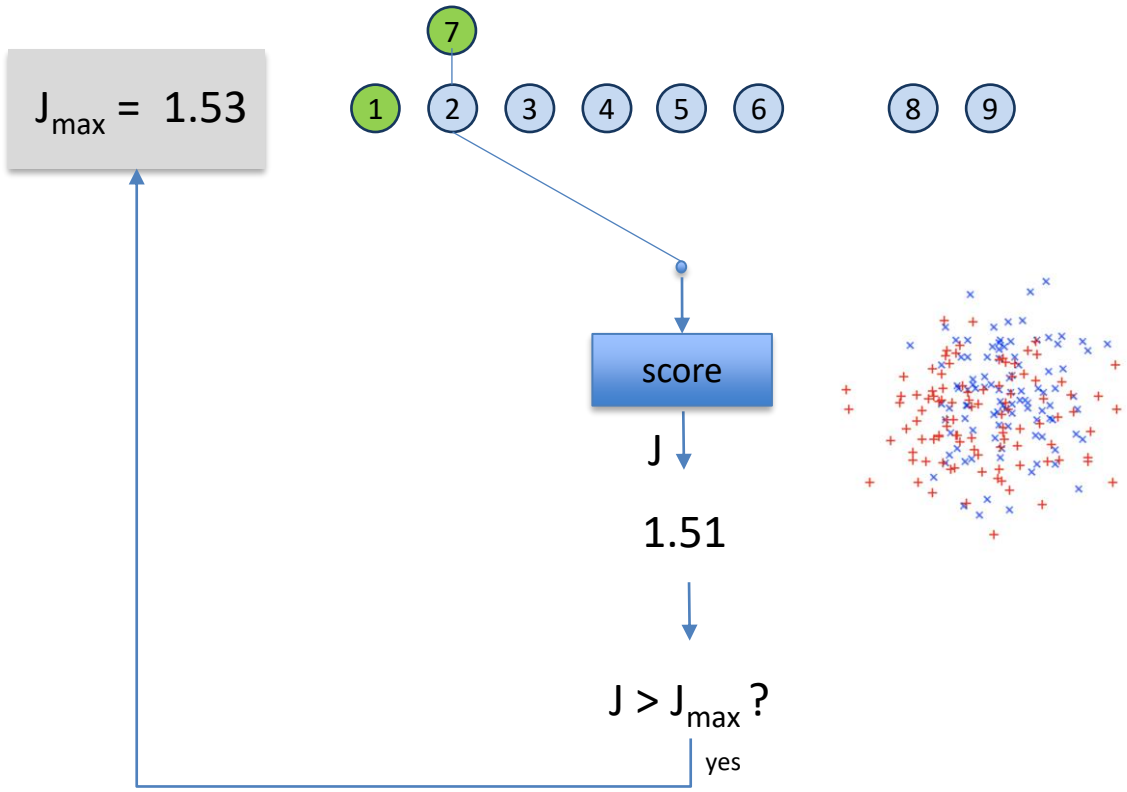

J_{max} = 1.47
1
2
3
4
5
6
7
8
9
score
J
0.27
J > J_{max} ?
yes

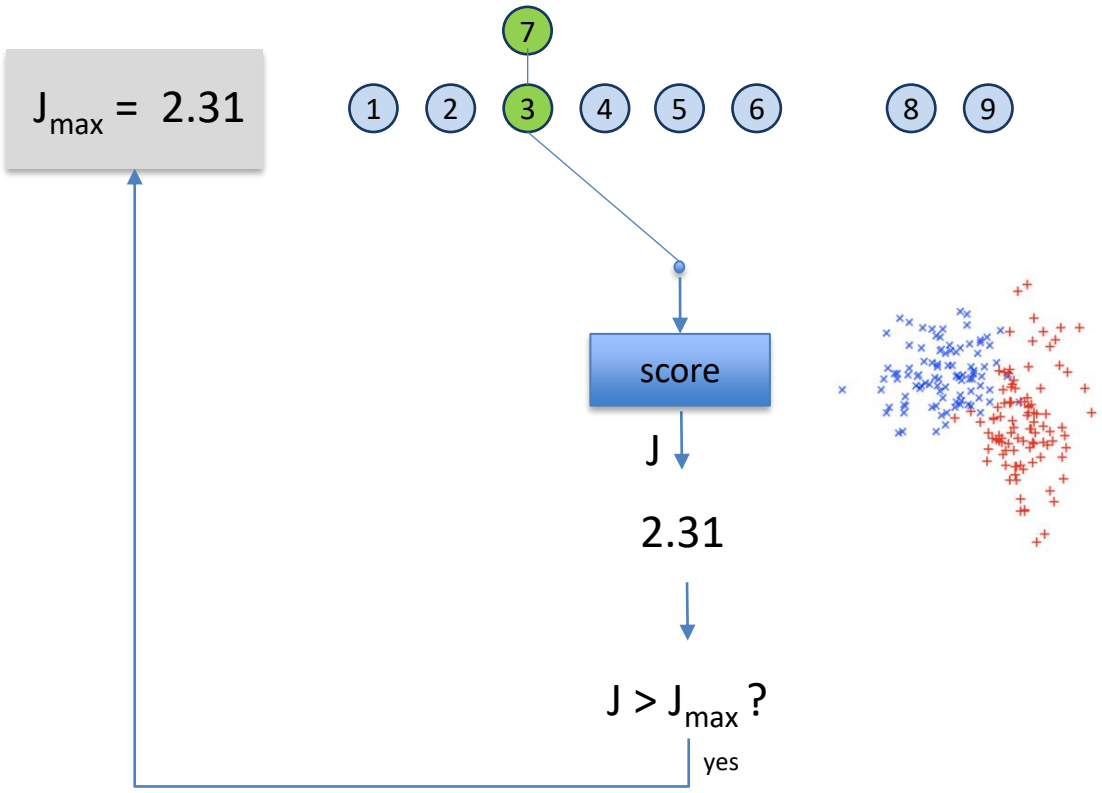

J_{max} = 1.47
1
2
3
4
5
6
7
8
9
score
J
0.13
J > J_{max} ?
yes

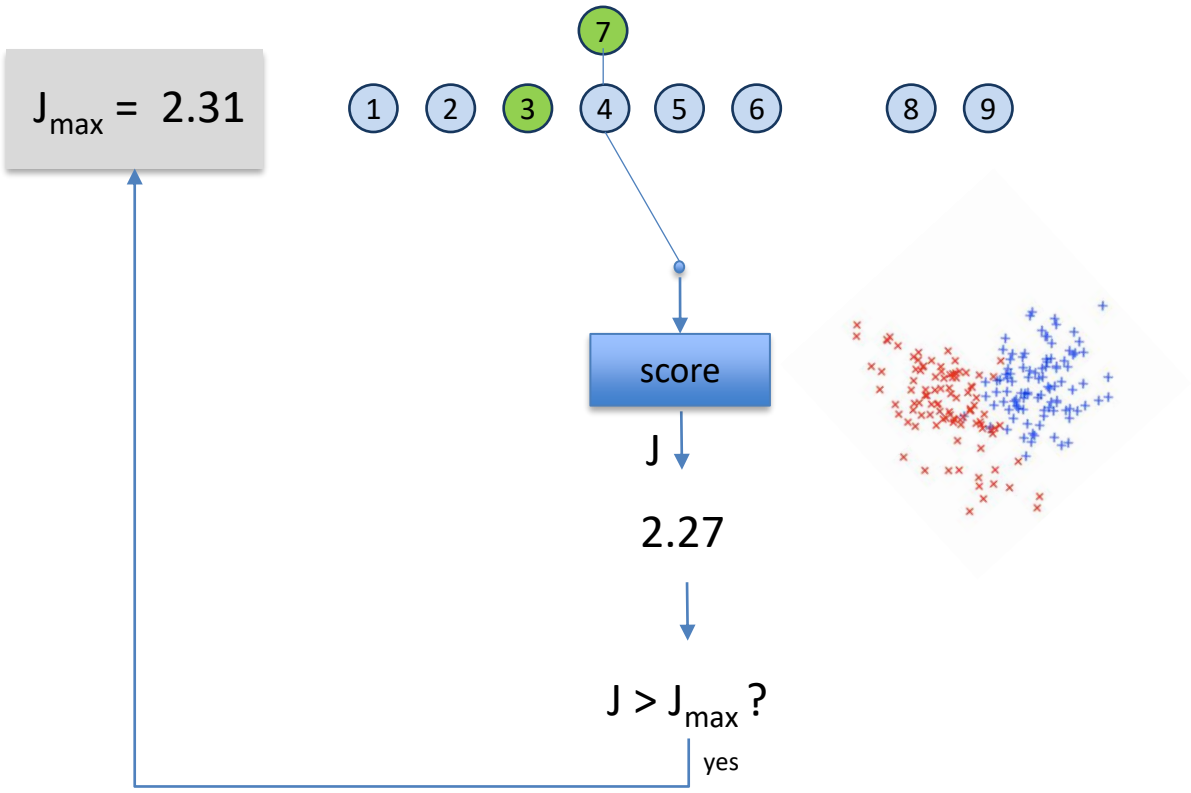
$J_{max} = 1.47$

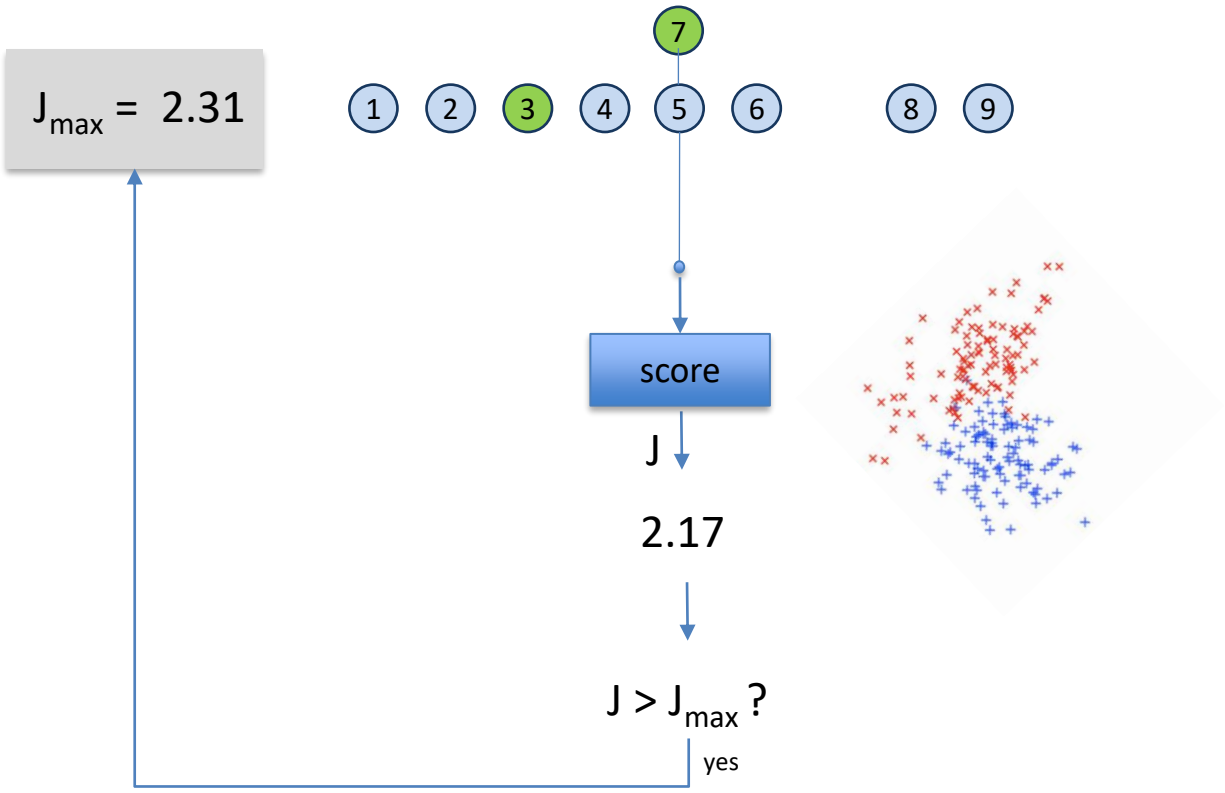


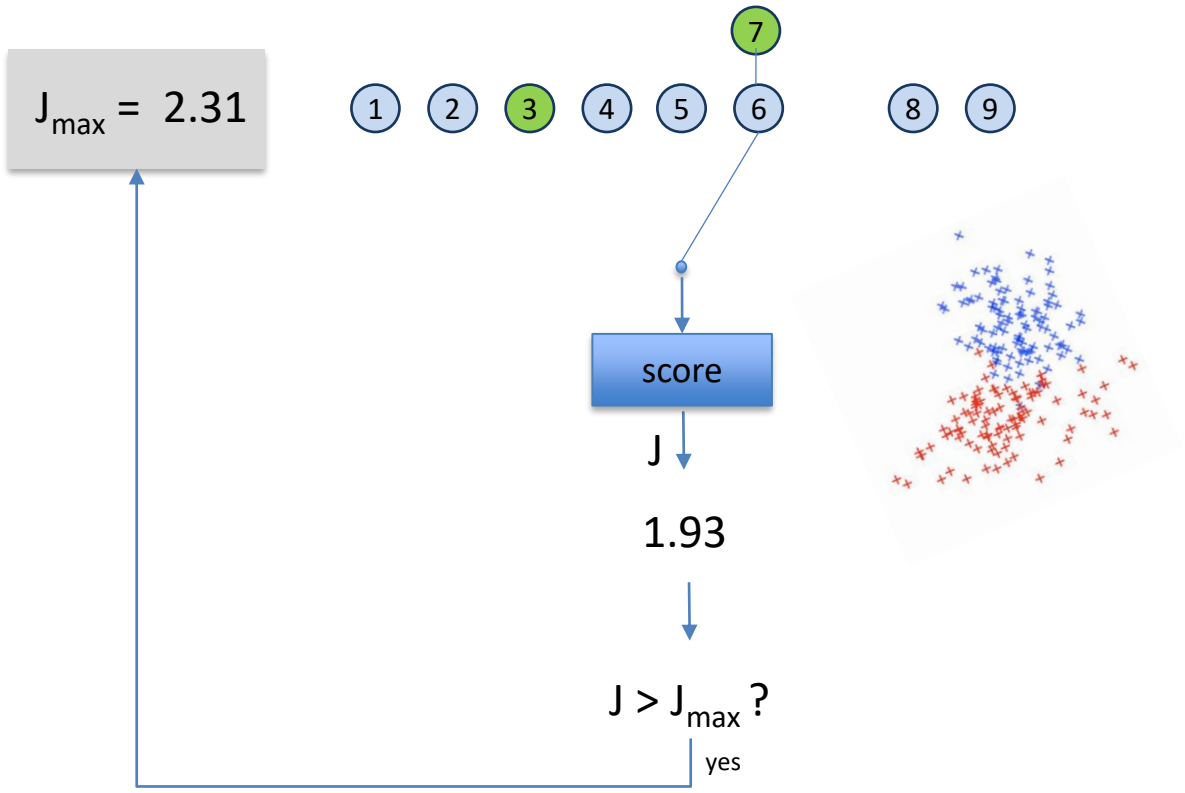

J_{max} = 1.53
7
1
2
3
4
5
6
8
9
score
J
1.53
$J > J_{max}$?
yes

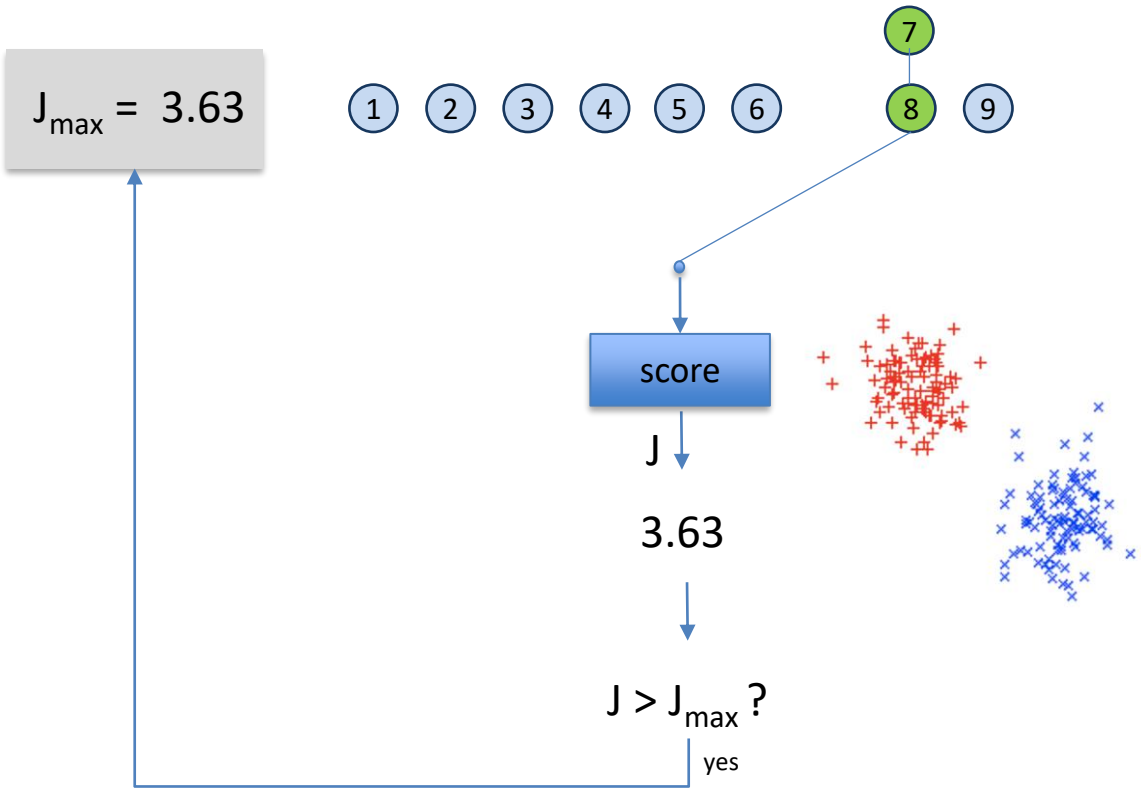

J_{max} = 1.53
7
1
2
3
4
5
6
8
9
score
J
1.51
$J > J_{max}$?
yes

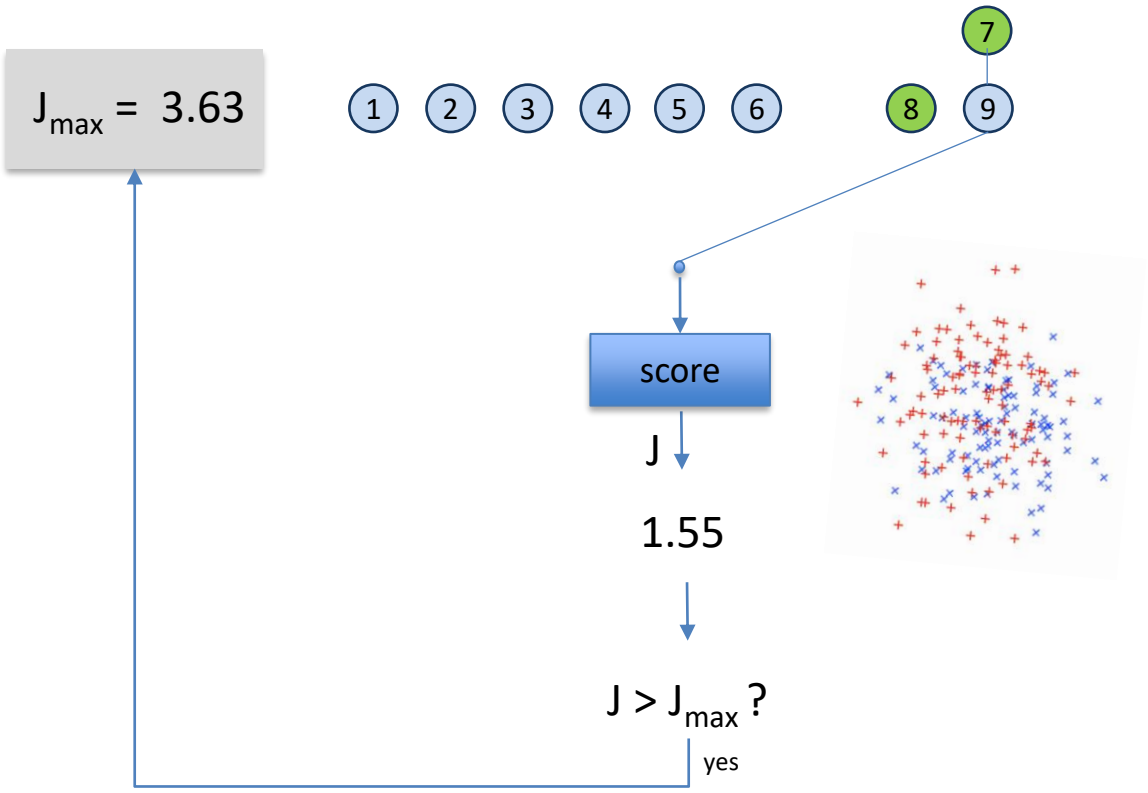

J_{max} = 2.31
7
1
2
3
4
5
6
8
9
score
J
2.31
$J > J_{max}$?
yes


J_{max} = 2.31
7
1
2
3
4
5
6
8
9
score
J
2.27
$J > J_{max}$?
yes

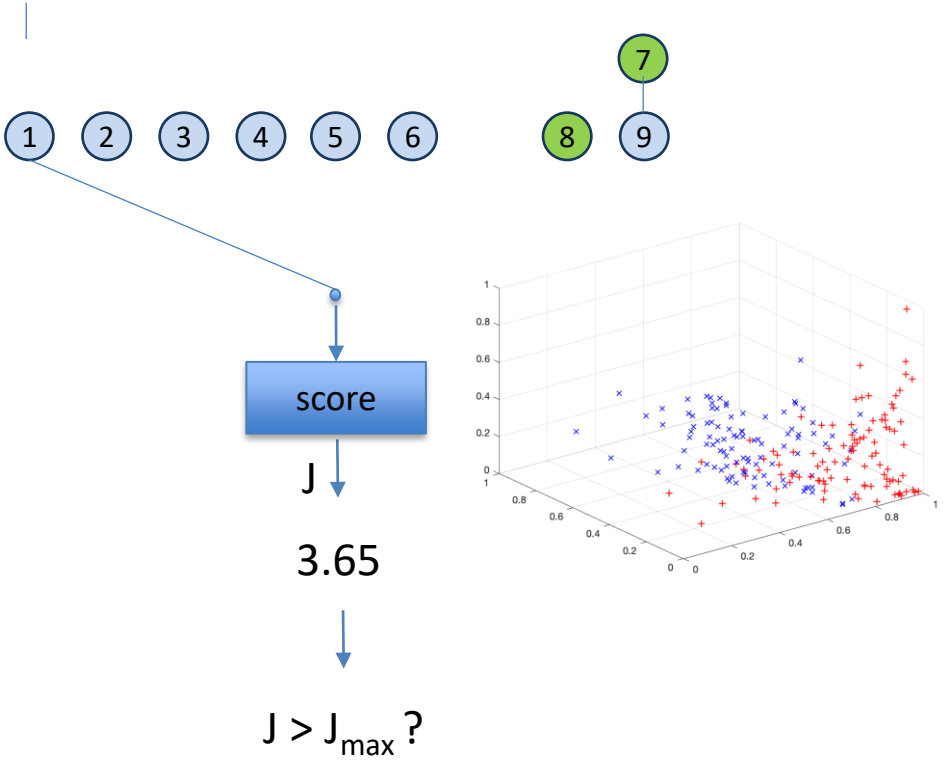

J_{max} = 2.31
1
2
3
4
5
6
7
8
9
score
J
2.17
$J > J_{max}$?
yes

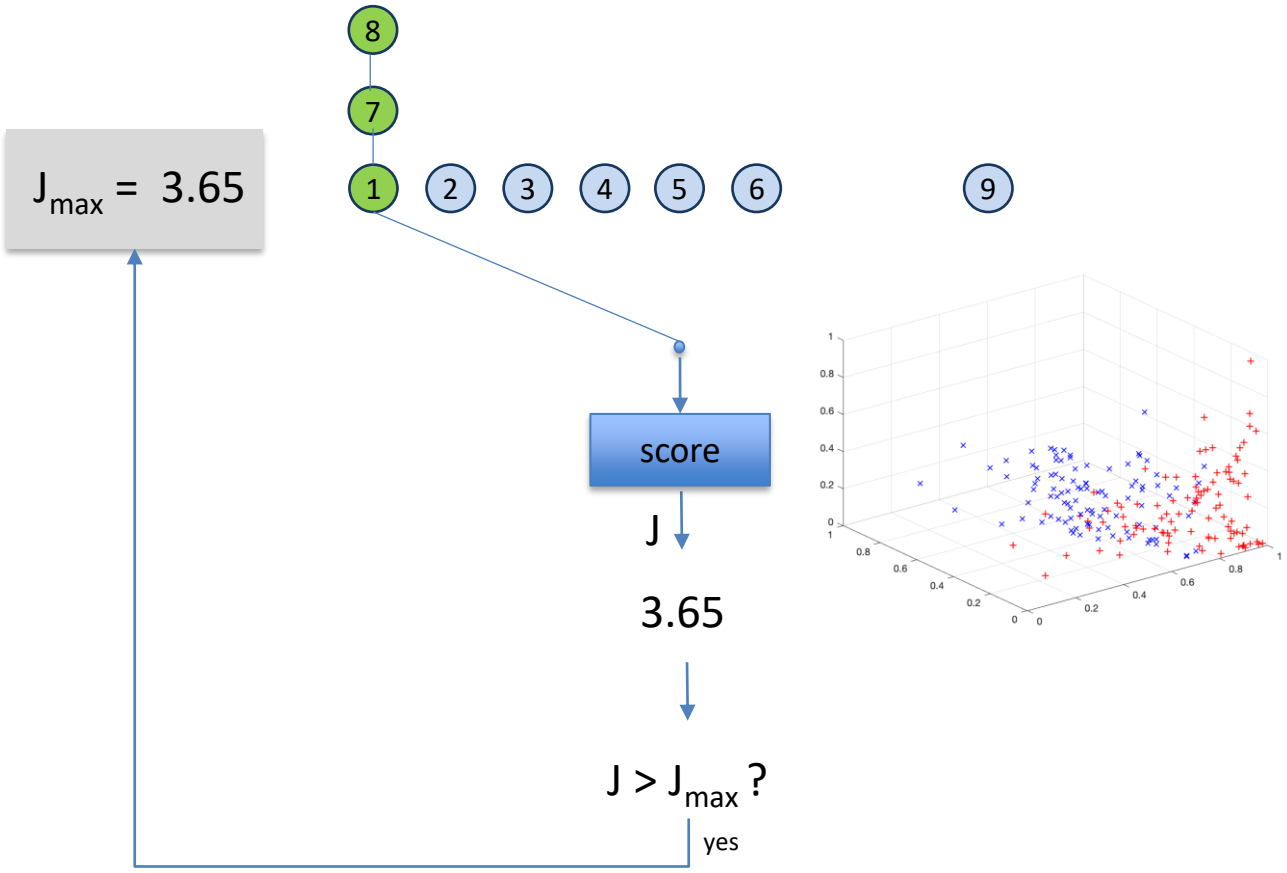

J_{max} = 2.31
1
2
3
4
5
6
7
8
9
score
J
1.93
J > J_{max} ?
yes

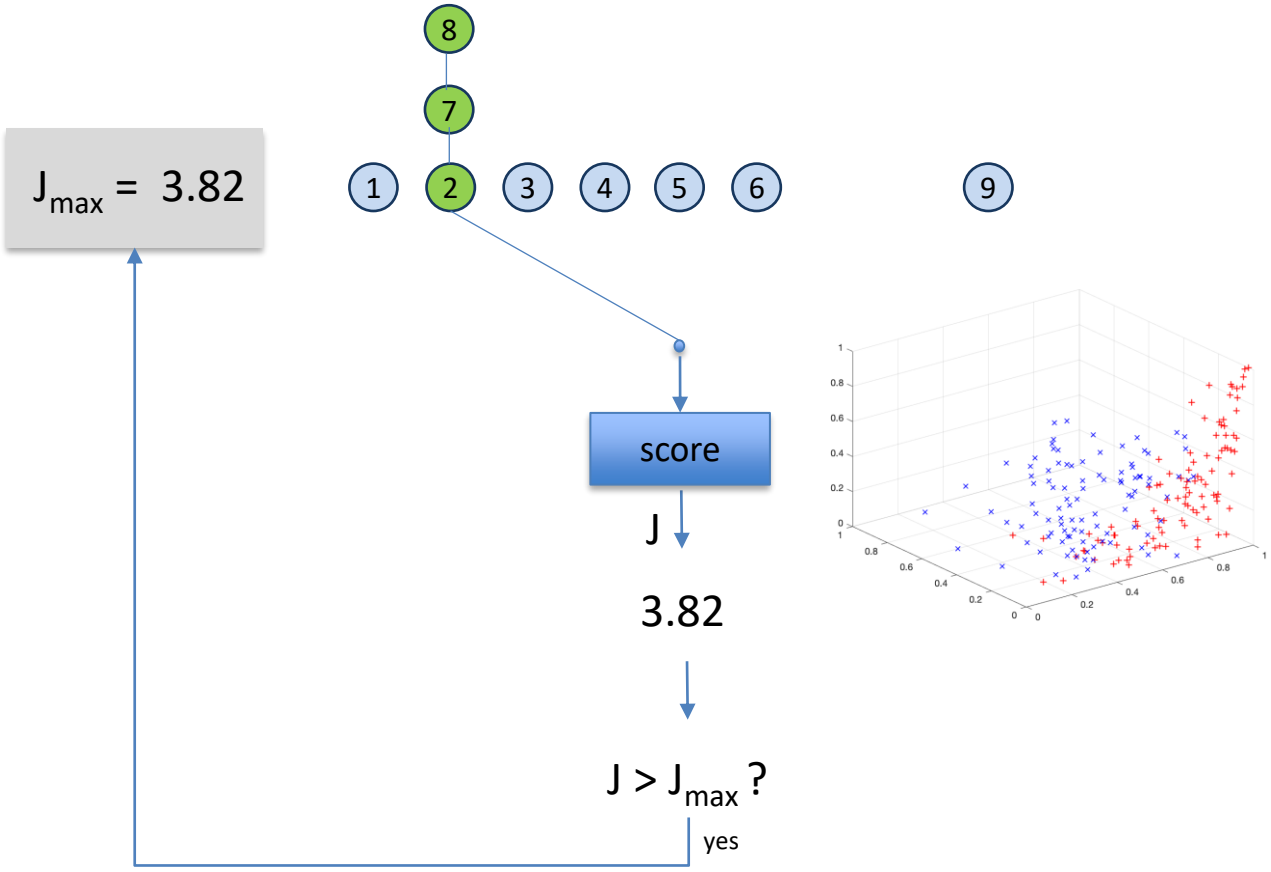

J_{max} = 3.63
1
2
3
4
5
6
7
8
9
score
J
3.63
J > J_{max} ?
yes

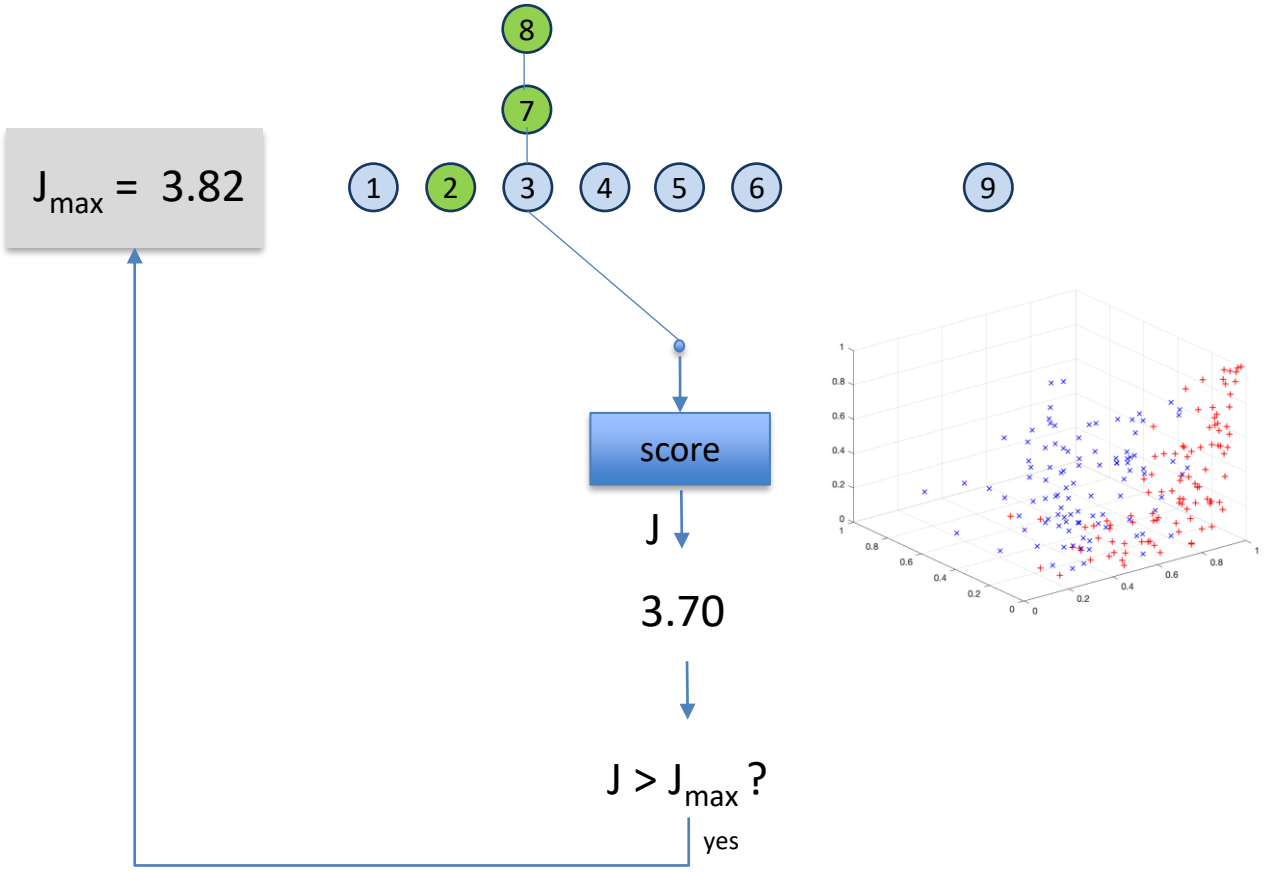

J_{max} = 3.63
1
2
3
4
5
6
7
8
9
score
J
1.55
$J > J_{max}$?
yes

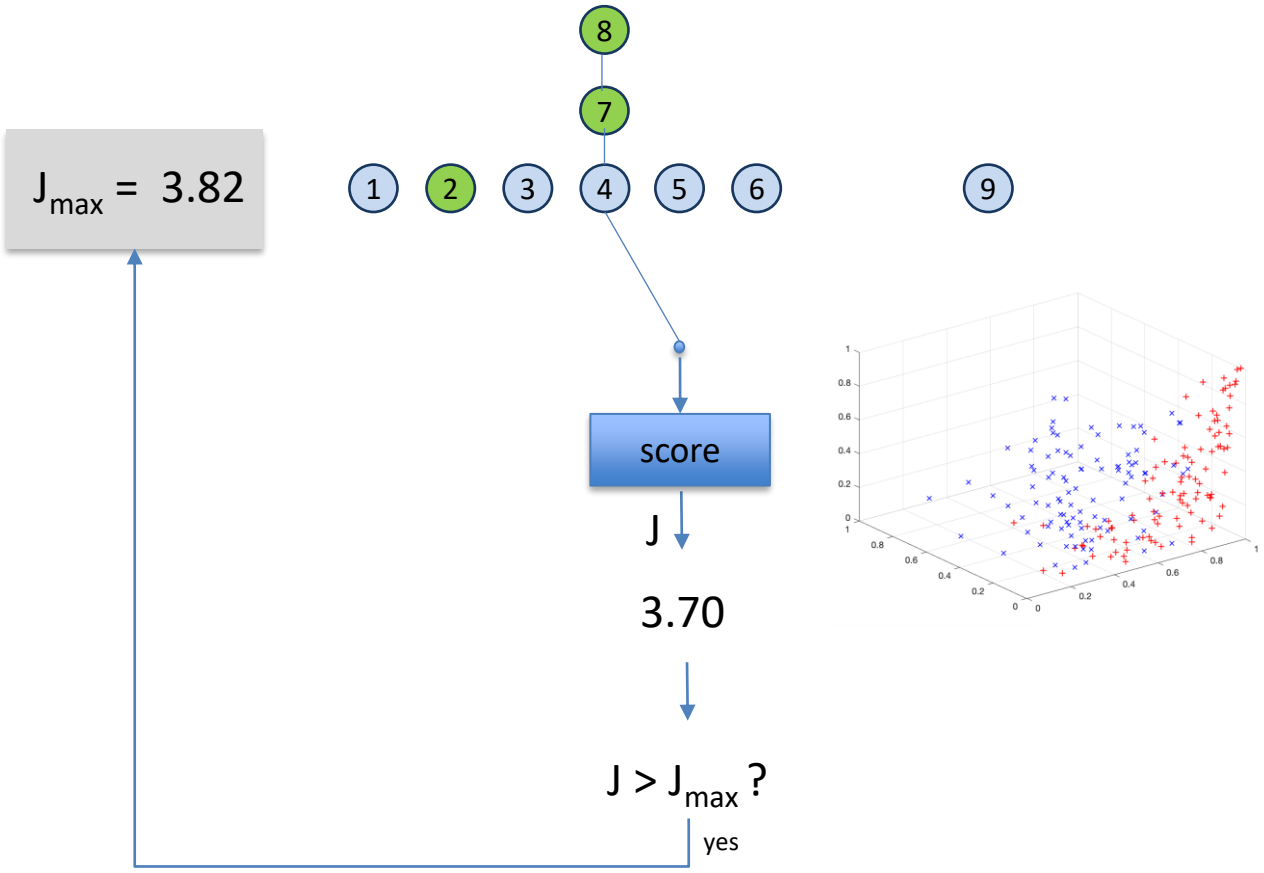
$J_{max} = 3.63$

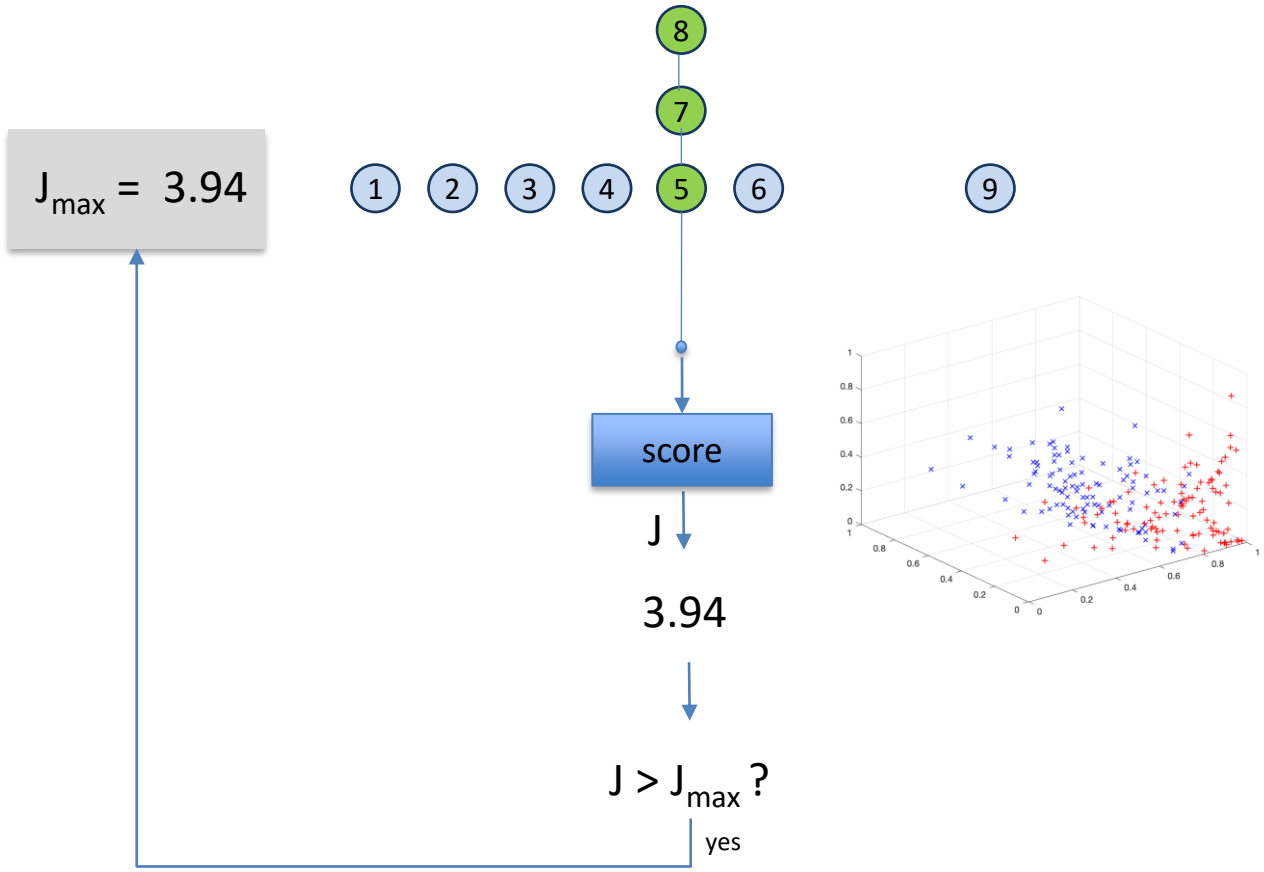


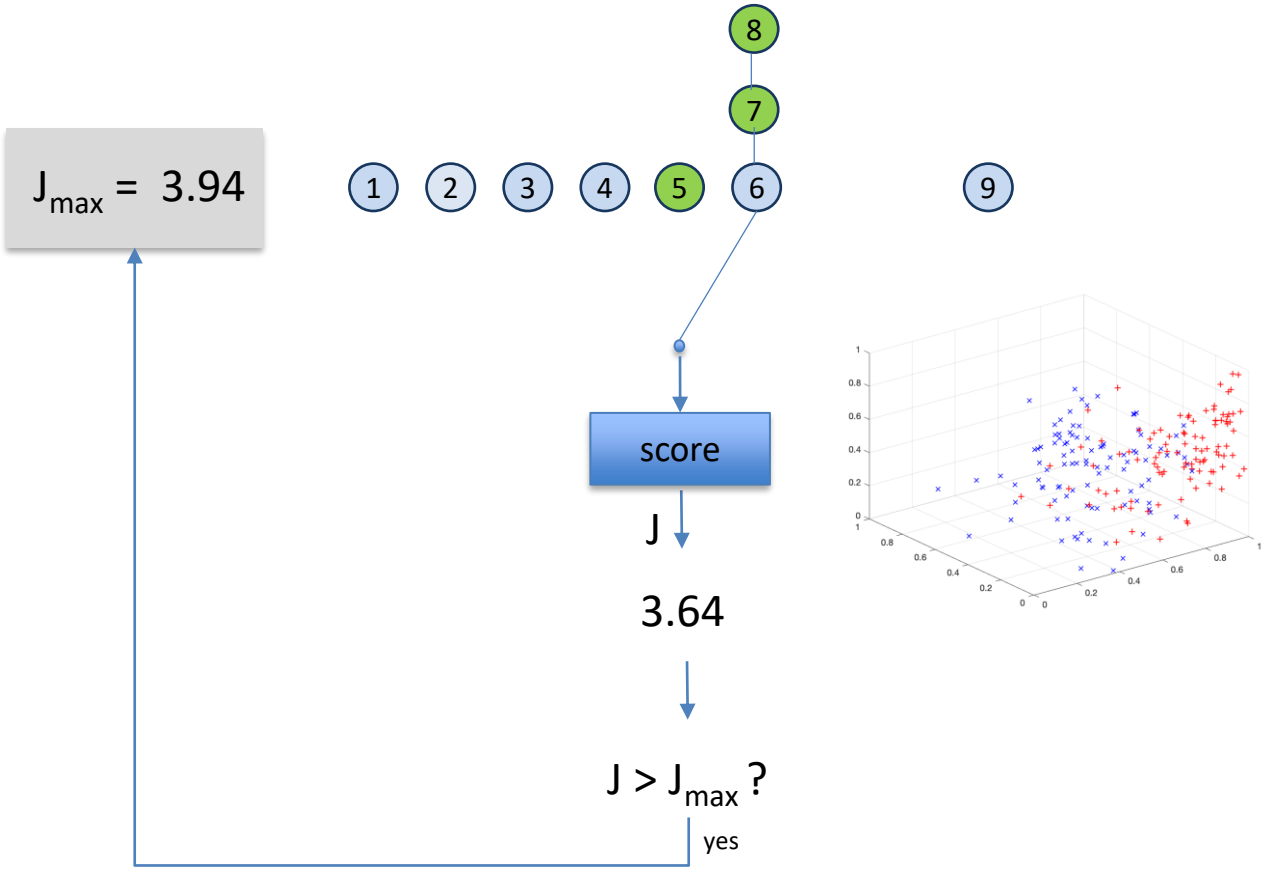

8
7
1
2
3
4
5
6
9
J_{max} = 3.65
score
J
3.65
$J > J_{max}$?
yes

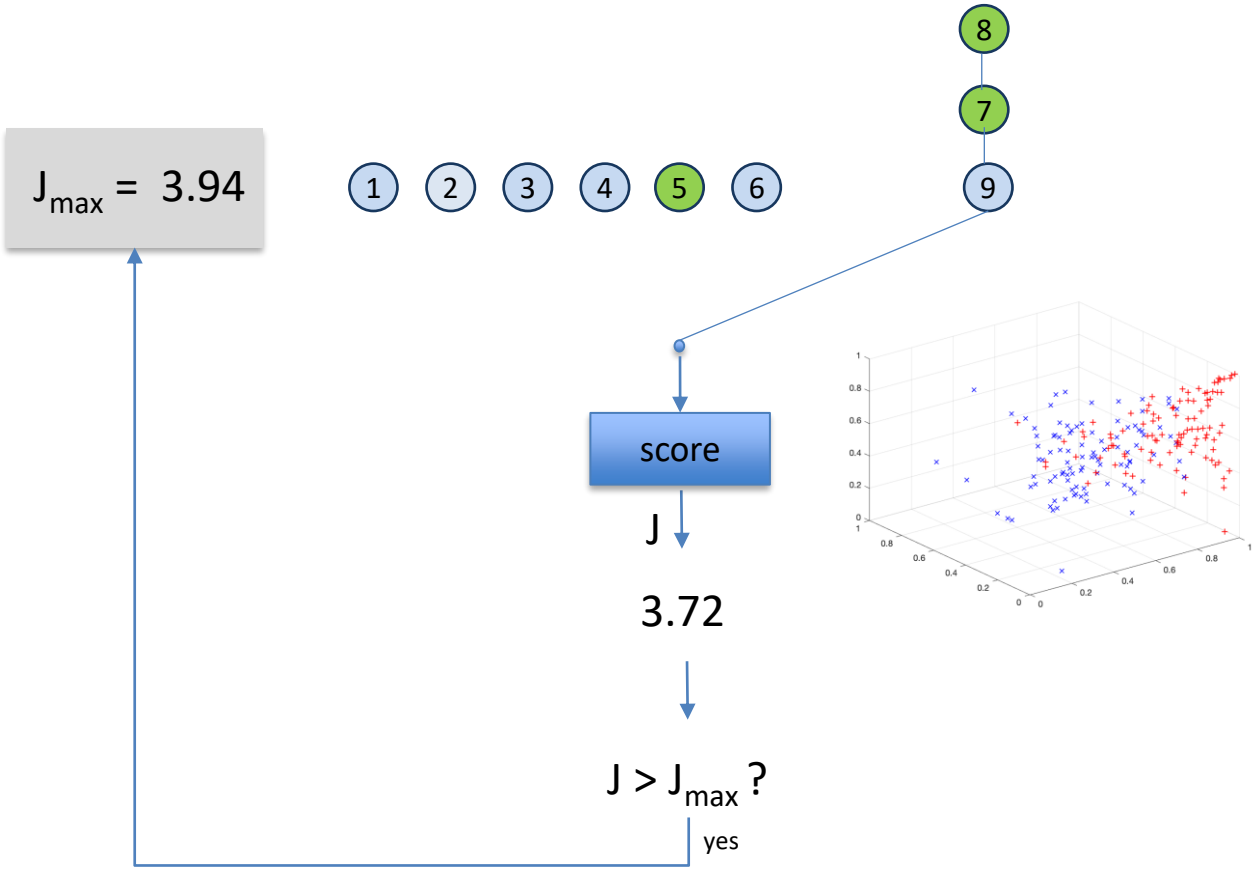

8
7
1
2
3
4
5
6
9
J_{max} = 3.82
score
J
3.82
$J > J_{max}$?
yes


8
7
1
2
3
4
5
6
9
J_{max} = 3.82
score
J
3.70
$J > J_{max}$?
yes


8
7
J_{max} = 3.82
1
2
3
4
5
6
9
score
J
3.70
$J > J_{max}$?
yes

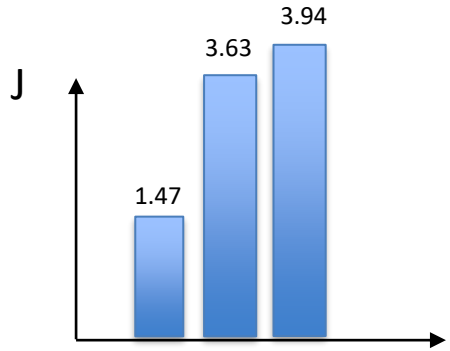

8
7
1
2
3
4
5
6
9
J_{max} = 3.94
score
J
3.94
$J > J_{max}$?
yes


8
7
J_{max} = 3.94
1
2
3
4
5
6
9
score
J
3.64
J > J_{max} ?
yes


J_{max} = 3.94
1
2
3
4
5
6
7
8
9
score
J
3.72
$J > J_{max}$?
yes

$J_{max} = 3.94$

1 2 3 4 5 6 7 8 9



Selected Features

SFS Algorithm

```
s    = [ ]                      # selected features
a    = ones(m,1)                # available features
Jmax = 0                        # initialization
for j = 1 to p                  # p features to be selected
    for k = 1 to m              # X has m features
        if a(k)  ==  1          # is feature k available
            st = [s, k]         # columns of X
            J = score(X(st))    # evaluation of the score
            if J>Jmax
                kmax = k
                Jmax = J        # update maximum
            end
        end
    end
    a(kmax) = 0                 # selected features in no available
    s = [s, kmax]               # update selected features
end
```